Special Publication ARFSD-SP-94001

BRIEF TUTORIAL ON THE KALMAN FILTER

John Podesta

November 1994

US ARMY
TANK AUTOMOTIVE AND
ARMAMENTS COMMAND
ARMAMENT RDE CENTER

U.S. ARMY ARMAMENT RESEARCH, DEVELOPMENT AND ENGINEERING CENTER

Fire Support Armaments Directorate

Picatinny Arsenal, New Jersey

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE November 1994	3. REPORT TYPE AND DATES COVERED Annual/Sep 91 to Nov 91

4. TITLE AND SUBTITLE BRIEF TUTORIAL ON THE KALMAN FILTER	5. FUNDING NUMBERS
6. AUTHOR(S) John Podesta	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESSES(S) ARDEC, FSAC Precision Munitions Division (SMCAR-FSP-E) Picatinny Arsenal, NJ 07806-5000	8. PERFORMING ORGANIZATION REPORT NUMBER Special Publication ARFSD-SP-94001
9.SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(S) ARDEC, IMD Information Research Center (AMSTA-AR-IMC) Picatinny Arsenal, NJ 07806-5000	10. SPONSORING/MONITORING AGENCY REPORT NUMBER
11. SUPPLEMENTARY NOTES	
12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.	12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

A derivation of the Kalman filter equations is presented which should provde a concise introduction to Kalman filter theory for scientists, engineers, and mathematicians alike. An elementary derivation of the basic Kalman filter, the 1-step Kalman predictor, is given first in 1-dimension and then in n-dimension. The well known prediction-correction formulation of the Kalman filter equations are derived for the filtered estimate, or current state estimate. Then, the state prediction is obtianed from the filtered estimate. It is assumed that the reader has a background in probability theory and some exposure to stochastic processes.

14. SUBJECT TERMS				15. NUMBER OF PAGES 31
Kalman filter	Kalman filtering	Linear filtering	Estimation theory	
Recursive estimation	Least squares estimation	Linear prediction theory		16. PRICE CODE
Stochastic processes	Stochastic control			

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR

Contents

1 Introduction

Since its invention by Rudolf E. Kalman in 1960, (refs. 1 through 5)[1]–[5] the Kalman filter has found widespread applications in commercial industry, the defense industry, and in academia. (In this report, all references are enclosed in square brackets, i.e., [1] stands for reference 1, etc.; and all equations are indicated by equation numbers in parentheses, i.e., (1) stands for equation 1, etc.) The Kalman filter is especially important in military applications where guidance, control, and signal processing systems must achieve a high level of performance. Although many books and articles have been published on the theory of the Kalman filter [6]–[11] the derivation of the Kalman filter equations is often difficult to follow. Unfortunately, this may leave the unacquainted engineer or scientist without quick access to the main results. For many, there is little time to wade through all the mathematical proofs and derivations which lead, finally, to the Kalman filter equations. Nevertheless, the derivation of the filter equations is not difficult, and for educational purposes it is desireable to simplify it as much as possible. The purpose of this brief report is to present a straightforward derivation of the Kalman filter which is concise and simple. A simple derivation of the filter equations should make the theory readily accessible to anyone who is interested.

It is assumed that the reader has achieved a certain level of mathematical maturity which includes a good knowledge of calculus, linear algebra, and probability theory. Also, the reader should be familiar with linear systems theory and the state space description of linear systems. Furthermore, it is desireable, though not essential, that the reader has had an introduction to stochastic processes. For those who are not familiar with stochastic processes, section 3 may be safely omitted without loss of continuity. Throughout this report all vectors and matrices will be written in boldface type, and it is assumed that all vectors and matrices are real valued. Regarding terminology, the terms random variable and random vector will be used interchangeably since this should not lead to confusion.

The report is structured as follows. The method of least squares is reviewed in section number 2. In section 3 it is shown that for statistical reasons the least squares estimate of a random quantity must be considered to be a random variable. The Kalman filter equations are derived in sections 4, 5, 6, and 7. In section 4, the basic Kalman filter, or Kalman predictor, is derived in 1-dimension. This is then extended to the $n-$dimensional case in section 5. In section 6, the standard Kalman filter equations are derived in the general $n-$dimensional case. The initial conditions for the standard Kalman filter require some special consideration, and these are derived in section 7. Relations between the Kalman predictor and the standard Kalman filter are developed in section 8.

2 Least Squares Estimation

Consider a physical system in a steady state (not time varying) and let the state be described by a state variable x. In the steady state, x is a constant. Suppose that a measurement made on the system yields a value z. The measurement process is not perfect

and will inevitably contain errors, i.e., experimental errors. Thus the measurement process can be modeled by the measurement equation

$$z = Hx + v, \tag{1}$$

where H is a scale factor which is built into the measurement equipment and v is a random variable which characterizes the measurement errors. The probability density function (PDF) of the random variable v is assumed given. Typically, v has a normal distribution with zero mean and known variance. Once a measurement z has been obtained in the lab, the problem is to find an estimate $\hat{x}$ of the state x based on the noise-corrupted measurement z. It is intuitively obvious that if the error is small then the desired estimate is simply z/H. In mathematical terms, this problem can be solved using the method of least squares. Let $\hat{x}$ be an unknown variable and define the error

$$e = z - H\hat{x}. \tag{2}$$

Since z and H are fixed real numbers, the error is a function of the variable $\hat{x}$. In the method of least squares, the best estimate $\hat{x}$ is that which minimizes the square error

$$\varepsilon = (z - H\hat{x})^2. \tag{3}$$

Since $\varepsilon = \varepsilon(\hat{x})$ is a function of $\hat{x}$, this will be a minimum if

$$\frac{d\varepsilon}{d\hat{x}} = 0. \tag{4}$$

Differentiating (3) yields

$$\frac{d\varepsilon}{d\hat{x}} = 2(z - H\hat{x})(-H) = 0. \tag{5}$$

Hence, the solution is

$$\hat{x} = \frac{z}{H}, \tag{6}$$

as was expected. To verify that this is a minimum and not a maximum examine the second derivative

$$\frac{d^2\varepsilon}{d\hat{x}^2} = 2H^2 > 0. \tag{7}$$

Since this is positive, the value of $\hat{x}$ given by (6) is indeed a minimum.

Next, consider the same problem in n-dimensions. The generalized measurement model is given by

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{v}, \tag{8}$$

where $\mathbf{x}$ is an n-dimensional column vector, $\mathbf{z}$ is an m-dimensional column vector of measurements, and $\mathbf{H}$ is an $m \times n$ matrix. The m-dimensional column vector $\mathbf{v}$ is a zero mean random vector that describes the measurement errors. The PDF of $\mathbf{v}$ is assumed given. For simplicity, it will be assumed that $m = n$, and that $\mathbf{H}$ is nonsingular. Observe that the quantities $\mathbf{z}$ and $\mathbf{H}$ are given and the state vector $\mathbf{x}$ is to be estimated from the

measurements. Thus, the problem is to find the estimate $\hat{\mathbf{x}}$ of the state vector $\mathbf{x}$ which minimizes the square error

$$\varepsilon = ||\mathbf{z} - \mathbf{H}\hat{\mathbf{x}}||^2. \tag{9}$$

The square error is a scalar and may be written

$$\varepsilon = (\mathbf{z} - \mathbf{H}\hat{\mathbf{x}})^T(\mathbf{z} - \mathbf{H}\hat{\mathbf{x}}), \tag{10}$$

where the superscript T denotes the transpose. Expanding the product this becomes

$$\varepsilon = \mathbf{z}^T\mathbf{z} - \mathbf{z}^T\mathbf{H}\hat{\mathbf{x}} - \hat{\mathbf{x}}^T\mathbf{H}^T\mathbf{z} + \hat{\mathbf{x}}^T\mathbf{H}^T\mathbf{H}\hat{\mathbf{x}}. \tag{11}$$

All quantities are known except for $\hat{\mathbf{x}}$. Therefore, the square error ε is a function of the components of $\hat{\mathbf{x}}$, that is,

$$\varepsilon = f(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n). \tag{12}$$

Using a well known result from advanced calculus, this function will be a minimum if

$$\frac{\partial f}{\partial \hat{x}_i} = 0, \tag{13}$$

for $i = 1, 2, \ldots, n$. It is convenient to define the derivative of a scalar with respect to a vector to be the column vector with components

$$\left(\frac{\partial f}{\partial \hat{\mathbf{x}}}\right)_i = \frac{\partial f}{\partial \hat{x}_i}, \tag{14}$$

where

$$(\hat{\mathbf{x}})_i = \hat{x}_i, \tag{15}$$

$i = 1, 2, \ldots, n$. Thus, the square error will be a minimum if

$$\frac{\partial \varepsilon}{\partial \hat{\mathbf{x}}} = 0. \tag{16}$$

The terms in equation (11) can be differentiated by using the following identities:

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{y}^T\mathbf{x}) = \mathbf{y}, \tag{17}$$

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}^T\mathbf{y}) = \mathbf{y}, \tag{18}$$

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}^T\mathbf{A}\mathbf{x}) = 2\mathbf{A}\mathbf{x}, \tag{19}$$

where $\mathbf{x}$ and $\mathbf{y}$ are arbitrary vectors and $\mathbf{A}$ is an arbitrary matrix. These relations are easy to prove by writting them out in component form. Proceeding with the differentiation of equation (11), it follows that

$$\frac{\partial \varepsilon}{\partial \hat{\mathbf{x}}} = -\mathbf{H}^T\mathbf{z} - \mathbf{H}^T\mathbf{z} + 2\mathbf{H}^T\mathbf{H}\hat{\mathbf{x}} = 0. \tag{20}$$

Rearranging terms, this becomes

$$\mathbf{H}^T\mathbf{H}\hat{\mathbf{x}} = \mathbf{H}^T\mathbf{z}. \tag{21}$$

This system of linear equations are called the *normal equations*. Clearly, $\hat{\mathbf{x}}$ will minimize the square error (9) if it is a solution of the normal equations (21). If $\mathbf{H}$ is nonsingular, i.e., invertible, then the solution is given by

$$\hat{\mathbf{x}} = \mathbf{H}^{-1}\mathbf{z}. \tag{22}$$

This is the least squares estimate of the state vector $\mathbf{x}$ based on the noise-corrupted vector of measurements $\mathbf{z}$. Note the close similarity with the 1-dimensional solution (6). In the general case when m and n are arbitrary, the solution to the normal equations is given by

$$\hat{\mathbf{x}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{z}, \tag{23}$$

provided that the indicated inverse exists.

3 Prediction Problem

In the derivation of the Kalman filter equations in section 4, the estimate $\hat{\mathbf{x}}$ is considered to be a random variable and not a numerical vector as it was in section 2. The reason for this difference of interpretation will be explained in this section. For simplicity, only scalar (1-dimensional) variables will be considered here.

The linear prediction of stochastic processes involves the following problem. Given a realization of a random process with the observed values $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_N, \ldots$, find an estimate of the value $\hat{\mathbf{x}}_{N+1}$ of the process at time $N+1$ based on the preceeding values up to and including $\mathbf{x}_N$. This is called the prediction problem. The estimate is called linear if it is a linear function of the variables $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_N$, that is, if the estimate has the linear form

$$\hat{\mathbf{x}}_{N+1} = \sum_{k=0}^{N} a_k \mathbf{x}_k. \tag{24}$$

It is possible to solve this problem using the method of least squares by minimizing the squared error

$$\varepsilon = \|\hat{\mathbf{x}}_{N+1} - \mathbf{x}_{N+1}\|^2. \tag{25}$$

By inserting the expression (24) into equation (25), the coefficients $a_0, a_1, \ldots, a_N$ may be computed using the method of the previous section. However, the solution for the prediction coefficients $a_0, a_1, \ldots, a_N$ will be different for different realizations of the same process. That is to say that if $\mathbf{x}'_0, \mathbf{x}'_1, \ldots, \mathbf{x}'_N$ were another realization of the same stochastic process, then the resulting solution for the a_k will be different. It is desireable to choose the coefficients so that the same set of coefficients may be used for any arbitrary realization of the process $\mathbf{x}(n)$. That is, so that in a statistical sense the same coefficients work equally well for any realization of the stochastic process. To accomplish this, it is necessary to rephrase the problem as follows: Given the random variables $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_N$, find a random variable

$\hat{\mathbf{x}}_{N+1}$ which is a linear combination of the random variables $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_N$, such that the *mean square error*

$$\varepsilon = E[\|\hat{\mathbf{x}}_{N+1} - \mathbf{x}_{N+1}\|^2] \tag{26}$$

is a minimum. In this case, the resulting estimate is a random variable $\hat{\mathbf{x}}_{N+1}$ which approximates the random variable $\mathbf{x}_{N+1}$ in the sense that the variance of the error $\mathbf{e} = (\hat{\mathbf{x}} - \mathbf{x})$ is a minimum. This is called the *least squares estimate* or *minimum variance estimate* of the random variable $\mathbf{x}_{N+1}$.

Problem Statement: Let $x(n)$ be a weakly stationary stochastic process and assume that $E[x(n)] = 0$. Fix N. The problem is to determine the coefficients $a_1, a_2, \ldots, a_N$, so that the random variable defined by

$$y = \sum_{k=1}^{N} a_k x_{N-k}, \tag{27}$$

minimizes the mean square error

$$\varepsilon = E[\|x_N - y\|^2]. \tag{28}$$

This problem was first considered by the Russian mathematician Kolmogorov in 1938. A translation of Kolmogorov's original paper can be found in reference [12]. The solution is so straightforward that Kolmogorov did not even bother to write it down.

Problem Solution: The mean square error is given by

$$\varepsilon = E[x_N^2 - 2x_N y + y^2]. \tag{29}$$

Substituting the expression (27) for y this becomes

$$\varepsilon = E[x_N^2 - 2\sum_{k=1}^{N} a_k x_N x_{N-k} + \sum_{j=1}^{N}\sum_{k=1}^{N} a_j a_k x_{N-j} x_{N-k}]. \tag{30}$$

By the definition of the autocorrelation function $R(n, m)$, which for a stationary process takes the form

$$E[x_n x_m] = R(n - m), \tag{31}$$

the mean square error (30) can be written

$$\varepsilon = R(0) - 2\sum_{k=1}^{N} a_k R(k) + \sum_{j=1}^{N}\sum_{k=1}^{N} a_j a_k R(j - k). \tag{32}$$

Since the autocorrelation function is known, ε can be considered to be a function of the variables $a_1, a_2, \ldots, a_N$, that is,

$$\varepsilon = f(a_1, a_2, \ldots, a_N). \tag{33}$$

Hence, the values of the coefficients a_k which make $f(a_1, \ldots, a_N)$ a minimum must satisfy the equations

$$\frac{\partial \varepsilon}{\partial a_1} = 0, \tag{34}$$

$$\frac{\partial \varepsilon}{\partial a_2} = 0, \tag{35}$$

$$\vdots$$

$$\frac{\partial \varepsilon}{\partial a_N} = 0, \tag{36}$$

Differentiating equation (32) with respect to a_k it follows that

$$\frac{\partial \varepsilon}{\partial a_k} = -2R(k) + 2\sum_{j=1}^{N} a_j R(j-k) = 0, \tag{37}$$

where $k = 1, \ldots, N$. This yields the nonhomogeneous system of linear equations

$$\sum_{j=1}^{N} a_j R(j-k) = R(k), \tag{38}$$

where $k = 1, \ldots, N$. In matrix form, this may be written

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \tag{39}$$

where

$$\mathbf{A} = \begin{pmatrix} R(0) & R(1) & R(2) & R(3) & \cdots & R(N-1) \\ R(1) & R(0) & R(1) & R(2) & \cdots & R(N-2) \\ \vdots & & \ddots & & \ddots & \vdots \\ R(N-1) & \cdots & \cdots & \cdots & \cdots & R(0) \end{pmatrix}, \tag{40}$$

$$\mathbf{x} = \begin{pmatrix} a_1 \\ \vdots \\ a_N \end{pmatrix}, \quad \text{and} \quad \mathbf{b} = \begin{pmatrix} R_1 \\ \vdots \\ R_N \end{pmatrix}. \tag{41}$$

If the matrix $\mathbf{A}$ is nonsingular, then the system (39) has the solution

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}. \tag{42}$$

This gives the solution for the prediction coefficients $a_1, \ldots, a_N$. For a good introduction to the linear prediction theory of stochastic processes, the reader should consult the original edition of the book by Yaglom [13]

The solution of the prediction problem is not really important here. What is important is the idea involved in the formulation of the problem. The most important point of this section is that in a statistical sense the estimate $\hat{x}_N$ must be a random variable, not a number. In practice, if a particular (numerical) realization $x_0, x_1, \ldots, x_{N-1}$ is observed, then the estimation formula (27) may be used to obtain a numerical result. However, the estimation formula is a general result that applies for any realization of the stochastic process.

4 Kalman Filter in 1-Dimension

In this section, the Kalman filter equations will be derived for the 1-step Kalman predictor. Consider a dynamical system that evolves in time according to the state equations

$$x_{n+1} = F_n x_n + w_n, \tag{43}$$

$n = 0, 1, 2, \ldots$, where $\{F_n\}_{n=0}^{\infty}$ is a given sequence of real numbers and $\{w_n\}_{n=0}^{\infty}$ is an independent, identically distributed (i.i.d.) sequence of zero mean random variables. The random sequence $\{w_n\}$ is often refered to as a white noise sequence since any two members of the sequence are uncorrelated, i.e., $E[w_n w_m] = 0$ for $n \neq m$. The noise-corrupted measurements of the system state are described by the measurement equation

$$z_n = H_n x_n + v_n, \tag{44}$$

$n = 0, 1, 2, \ldots$, where H_n is a proportionality factor (usually a constant) which is built into the measuring device. In general, $\{H_n\}_{n=0}^{\infty}$ is a given sequence of real numbers. The measurement errors or measurement noise is described by the random sequence $\{v_n\}$ which is a zero mean i.i.d. sequence. It is further assumed that the sequences $\{w_n\}$ and $\{v_n\}$ are uncorrelated and that

$$\begin{aligned} E[v_n^2] &= \sigma_v^2, && (45)\\ E[w_n^2] &= \sigma_w^2, && (46)\\ E[v_n w_m] &= 0, && (47) \end{aligned}$$

where $n, m = 0, 1, 2, \ldots$, and where σ_v^2 and σ_w^2 are given real numbers. (In general, the variances of v_n and w_n may depend on n, and the results of this section are easily extended to include this case.) To complete the specification of the system dynamics it is necessary to give some initial condition for x_0. Instead of giving a fixed initial condition, x_0 will be a random variable with a known probability density function $f(x_0)$. Typically, x_0 is normally distributed with mean $\bar{x}_0$ and variance σ_0^2, that is

$$x_0 \sim N(\bar{x}_0, \sigma_0^2).$$

Furthermore, the random variable x_0 is assumed to be uncorrelated with the noise sequences $\{w_n\}$ and $\{v_n\}$. There is one remark which should be made about the initial condition: the PDF $f(x_0)$ is not needed in the derivation and it is, therefore, irrelevant; it is only necessary to know the mean $\bar{x}_0$ and the variance σ_0^2 of the random variable x_0. Now to the problem at hand. In simple terms, the purpose of the Kalman filter is to make a prediction of the future state x_{n+1} at time $n+1$ given noise-corrupted measurements $z_0, z_1, \ldots, z_n$ from the time $k = 0$ up to the time $k = n$.

Problem Statement: Given the measurements $z_0, z_1, \ldots, z_n$, find the unbiased least squares estimate $\hat{x}_{n+1}$ for the state variable x_{n+1}. Assume that the estimate has the linear recursive form

$$\hat{x}_{n+1} = K_n z_n + K'_n \hat{x}_n. \tag{48}$$

Note that the recursive form (48) implies that the estimate $\hat{x}_{n+1}$ is a *linear combination* of the random variables $z_0, z_1, \ldots, z_n$. Consequently, $\hat{x}_{n+1}$ is a random variable too, and it makes sense to discuss the expectation of $\hat{x}_{n+1}$. This will appear in the derivation.

Problem Solution: If the estimate $\hat{x}_{n+1}$ is unbiased, then by definition $E[\hat{x}_{n+1} - x_{n+1}] = 0$. Using equations (43) and (48) one has

$$\hat{x}_{n+1} - x_{n+1} = (K_n z_n + K_n' \hat{x}_n) - (F_n x_n + w_n). \tag{49}$$

Substituting equation (44) and rearranging terms this yields

$$\hat{x}_{n+1} - x_{n+1} = (K_n H_n - F_n) x_n + K_n' \hat{x}_n + K_n v_n - w_n. \tag{50}$$

Taking the expectation of both sides of this equation gives

$$E[\hat{x}_{n+1} - x_{n+1}] = (K_n H_n - F_n) E[x_n] + K_n' E[\hat{x}_n] = 0. \tag{51}$$

But by hypothesis $\hat{x}_n$ is unbiased, and therefore $E[x_n] = E[\hat{x}_n]$. Thus, equation (51) implies

$$K_n' = F_n - K_n H_n. \tag{52}$$

Having solved for K_n', it is now necessary to determine K_n and the derivation will be complete.

The coefficient K_n must be chosen to minimize the mean square error

$$\varepsilon = E[(\hat{x}_{n+1} - x_{n+1})^2]. \tag{53}$$

Using the result (52), equation (50) can be written

$$(\hat{x}_{n+1} - x_{n+1}) = (F_n - K_n H_n)(\hat{x}_n - x_n) + K_n v_n - w_n. \tag{54}$$

Squaring this equation and taking the expectation of both sides yields

$$E[(\hat{x}_{n+1} - x_{n+1})^2] = (F_n - K_n H_n)^2 E[(\hat{x}_n - x_n)^2] + K_n^2 \sigma_v^2 + \sigma_w^2. \tag{55}$$

It is useful to define the *error variance*

$$P_n = E[(\hat{x}_n - x_n)^2]. \tag{56}$$

The previous equation may then be written in the form

$$P_{n+1} = (F_n - K_n H_n)^2 P_n + K_n^2 \sigma_v^2 + \sigma_w^2. \tag{57}$$

This equation expresses the mean square error $\varepsilon = P_{n+1}$ as a function of K_n. From elementary calculus, this will be a minimum when

$$\frac{d\varepsilon}{dK_n} = 0.$$

This implies

$$\frac{dP_{n+1}}{dK_n} = 2(F_n - K_nH_n)(-H_n)P_n + 2K_n\sigma_v^2 = 0. \tag{58}$$

Solving for K_n, one finds

$$K_n = \frac{F_nH_nP_n}{H_n^2P_n + \sigma_v^2}. \tag{59}$$

To verify that this is a minimum and not a maximum take the second derivative

$$\frac{d^2P_{n+1}}{dK_n^2} = 2H_n^2P_n + 2\sigma_v^2 > 0.$$

Since H_n^2, P_n, and σ_v^2 are all positive, the second derivative is greater than zero which implies that the value of K_n given by (59) is indeed a minimum. Substituting the result (59) into equation (57) yields

$$P_{n+1} = \frac{F_n^2P_n\sigma_v^2}{H_n^2P_n + \sigma_v^2} + \sigma_w^2, \tag{60}$$

which may be written in the equivalent form

$$P_{n+1} = F_n(F_n - K_nH_n)P_n + \sigma_w^2. \tag{61}$$

This completes the derivation in the 1-dimensional case.

In summary, the Kalman filter equations in 1-dimension are given by

$$\hat{x}_{n+1} = F_n\hat{x}_n + K_n(z_n - H_n\hat{x}_n), \tag{62}$$

$$P_{n+1} = (F_n - K_nH_n)F_nP_n + \sigma_w^2, \tag{63}$$

$$K_n = \frac{F_nH_nP_n}{P_nH_n^2 + \sigma_v^2}, \tag{64}$$

where $n = 0, 1, 2, \ldots$. In an actual implementation of the Kalman filter, the input is the measurement sequence z_n and the output is the sequence of state predictions $\hat{x}_{n+1}$. In real-time applications, the input and output are sequences of real numbers. The output $\hat{x}_{n+1}$ is obtained from the input z_n through equations (62), (63), and (64). In order to initialize the Kalman filter algorithm, it is necessary to specify the initial conditions. The initial conditions are given by

$$\hat{x}_0 = \bar{x}_0, \tag{65}$$

$$P_0 = \sigma_0^2. \tag{66}$$

It is interesting that the sequences P_n and K_n are not dependent on the measurements. Therefore, the sequences P_n and K_n can be precomputed before any measurements are taken. Of course, these quantities can also be computed in real-time if desired. In conclusion, one implementation of the Kalman filter algorithm is as follows:

1. Set $n = 0$,

2. Initialize $\hat{x}_0 = \bar{x}_0$,

3. Initialize $P_0 = \sigma_0^2$,

4. Calculate K_n from equation (64),

5. Calculate $\hat{x}_{n+1}$ from (62),

6. Calculate P_{n+1} from (63),

7. Increment n and goto step 4.

5 Kalman Filter in n-Dimensions

In n-dimensions, the system is characterized by an n-dimensional state vector $\mathbf{x}$ which satisfies a state equation of the form

$$\mathbf{x}_{k+1} = \mathbf{F}_k \mathbf{x}_k + \mathbf{w}_k, \tag{67}$$

where $k = 0, 1, 2, \ldots$, $\{\mathbf{F}_k\}$ is a given sequence of $n \times n$ real matrices, and $\{\mathbf{w}_k\}$ is an i.i.d. sequence of zero mean random vectors. The measurements of the system outputs are related to the state vector by the measurement equation

$$\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k, \tag{68}$$

where $\mathbf{z}_k$ is an m-dimensional vector of measurements, $\{\mathbf{H}_k\}$ is a given sequence of $m \times n$ matrices, and $\{\mathbf{v}_k\}$ is an i.i.d. sequence of zero mean random vectors. To complete the specification of the sequences $\{\mathbf{v}_k\}$ and $\{\mathbf{w}_k\}$, it is necessary to know the covariance matrices

$$E[\mathbf{v}_k \mathbf{v}_k^T] = \mathbf{R}_k, \tag{69}$$

$$E[\mathbf{w}_k \mathbf{w}_k^T] = \mathbf{Q}_k, \tag{70}$$

where $k = 0, 1, 2, \ldots$. In addition, the sequences $\{\mathbf{v}_k\}$ and $\{\mathbf{w}_k\}$ are uncorrelated so that

$$E[\mathbf{v}_j \mathbf{w}_k^T] = 0, \tag{71}$$

where j and k are any non-negative integers. The sequences of matrices $\{\mathbf{R}_k\}$ and $\{\mathbf{Q}_k\}$ are assumed to be given. In addition, these are both symmetric and, by assumption, positive definite matrices for each k. Finally, the initial condition on the state equation is that $\mathbf{x}_0$ is a random vector with a probability density function $f(\mathbf{x}_0)$. As remarked in section 4, it is not strictly necessary to know the PDF of $\mathbf{x}_0$, but only the mean $\bar{\mathbf{x}}_0$ and covariance matrix $\bar{\mathbf{P}}_0 = E[(\mathbf{x}_0 - \bar{\mathbf{x}}_0)(\mathbf{x}_0 - \bar{\mathbf{x}}_0)^T]$. Usually, $\mathbf{x}_0$ has a multivariate normal distribution with known mean and covariance matrix. Furthermore, it is required that $\mathbf{x}_0$ be uncorrelated with the sequences $\{\mathbf{v}_k\}$ and $\{\mathbf{w}_k\}$. Before proceeding with the derivation of the Kalman filter equations, observe that the vectors $\mathbf{x}_n$ and $\mathbf{z}_n$, which are defined by equations (67) and (68), are both random vectors.

Problem Statement: Given the measurements $\mathbf{z}_0, \mathbf{z}_1, \ldots, \mathbf{z}_n$, find the unbiased least squares estimate $\hat{\mathbf{x}}_{n+1}$ for the state vector $\mathbf{x}_{n+1}$. Assume that the estimate has the linear recursive form

$$\hat{\mathbf{x}}_{n+1} = \mathbf{K}_n \mathbf{z}_n + \mathbf{K}'_n \hat{\mathbf{x}}_n. \tag{72}$$

As in the 1-dimensional case, the vector $\hat{\mathbf{x}}_{n+1}$ is random since it is a function of the random vectors $\mathbf{z}_0, \mathbf{z}_1, \ldots, \mathbf{z}_n$.

Problem Solution: Since the estimate $\hat{\mathbf{x}}_{n+1}$ is unbiased, this implies $E[\hat{\mathbf{x}}_{n+1} - \mathbf{x}_{n+1}] = 0$. From equations (67) and (72),

$$\hat{\mathbf{x}}_{n+1} - \mathbf{x}_{n+1} = (\mathbf{K}_n \mathbf{z}_n + \mathbf{K}'_n \hat{\mathbf{x}}_n) - (\mathbf{F}_n \mathbf{x}_n + \mathbf{w}_n). \tag{73}$$

Substituting equation (68) for $\mathbf{z}_n$ and rearranging terms this becomes

$$\hat{\mathbf{x}}_{n+1} - \mathbf{x}_{n+1} = (\mathbf{K}_n \mathbf{H}_n - \mathbf{F}_n)\mathbf{x}_n + \mathbf{K}'_n \hat{\mathbf{x}}_n + \mathbf{K}_n \mathbf{v}_n - \mathbf{w}_n. \tag{74}$$

Taking the expectation of both sides of this equation gives

$$E[\hat{\mathbf{x}}_{n+1} - \mathbf{x}_{n+1}] = (\mathbf{K}_n \mathbf{H}_n - \mathbf{F}_n)E[\mathbf{x}_n] + \mathbf{K}'_n E[\hat{\mathbf{x}}_n]. \tag{75}$$

But by hypothesis $\hat{\mathbf{x}}_n$ is unbiased, and therefore $E[\mathbf{x}_n] = E[\hat{\mathbf{x}}_n]$. Thus, equation (75) implies

$$\mathbf{K}'_n = \mathbf{F}_n - \mathbf{K}_n \mathbf{H}_n. \tag{76}$$

This is the solution for $\mathbf{K}'_n$.

Next, the coefficient $\mathbf{K}_n$ will be obtained by minimizing the mean square error. This is similar to the procedure used in the 1-dimensional case. The mean square error is given by

$$\varepsilon = E[\|\hat{\mathbf{x}}_{n+1} - \mathbf{x}_{n+1}\|^2] = E[(\hat{\mathbf{x}}_{n+1} - \mathbf{x}_{n+1})^T(\hat{\mathbf{x}}_{n+1} - \mathbf{x}_{n+1})]. \tag{77}$$

For the purpose of taking derivatives, it is convenient to express this in the form of a trace. Therefore, write

$$\varepsilon = \mathrm{Tr}\ E[(\hat{\mathbf{x}}_{n+1} - \mathbf{x}_{n+1})(\hat{\mathbf{x}}_{n+1} - \mathbf{x}_{n+1})^T]. \tag{78}$$

Now define a new quantity

$$\mathbf{P}_{n+1} = E[(\hat{\mathbf{x}}_{n+1} - \mathbf{x}_{n+1})(\hat{\mathbf{x}}_{n+1} - \mathbf{x}_{n+1})^T]. \tag{79}$$

This is called the *error covariance matrix*. Then the mean square error becomes

$$\varepsilon = \mathrm{Tr}\ (\mathbf{P}_{n+1}). \tag{80}$$

In order to evaluate this, start by using equations (74) and (76) to write

$$(\hat{\mathbf{x}}_{n+1} - \mathbf{x}_{n+1}) = (\mathbf{F}_n - \mathbf{K}_n \mathbf{H}_n)(\hat{\mathbf{x}}_n - \mathbf{x}_n) + (\mathbf{K}_n \mathbf{v}_n - \mathbf{w}_n). \tag{81}$$

Thus,

$$(\hat{\mathbf{x}}_{n+1} - \mathbf{x}_{n+1})(\hat{\mathbf{x}}_{n+1} - \mathbf{x}_{n+1})^T = [(\mathbf{F}_n - \mathbf{K}_n\mathbf{H}_n)(\hat{\mathbf{x}}_n - \mathbf{x}_n) + (\mathbf{K}_n\mathbf{v}_n - \mathbf{w}_n)]$$
$$\times [(\hat{\mathbf{x}}_n - \mathbf{x}_n)^T(\mathbf{F}_n^T - \mathbf{H}_n^T\mathbf{K}_n^T) + (\mathbf{v}_n^T\mathbf{K}_n^T - \mathbf{w}_n^T)]. \tag{82}$$

By expanding the product and then using the linearity of the expectation, this equation yields

$$\mathbf{P}_{n+1} = (\mathbf{F}_n - \mathbf{K}_n\mathbf{H}_n)\mathbf{P}_n(\mathbf{F}_n^T - \mathbf{H}_n^T\mathbf{K}_n^T) + \mathbf{K}_n\mathbf{R}_n\mathbf{K}_n^T + \mathbf{Q}_n, \tag{83}$$

where

$$\mathbf{P}_n = E[(\hat{\mathbf{x}}_n - \mathbf{x}_n)(\hat{\mathbf{x}}_n - \mathbf{x}_n)^T], \tag{84}$$
$$\mathbf{R}_n = E[\mathbf{v}_n\mathbf{v}_n^T], \tag{85}$$
$$\mathbf{Q}_n = E[\mathbf{w}_n\mathbf{w}_n^T]. \tag{86}$$

Of course, use has been made of the fact that $\hat{\mathbf{x}}_n$ and $\mathbf{x}_n$ are both uncorrelated with $\mathbf{v}_n$ and $\mathbf{w}_n$. By expanding the first term in equation (83) this may be written

$$\begin{aligned}\mathbf{P}_{n+1} &= \mathbf{F}_n\mathbf{P}_n\mathbf{F}_n^T - \mathbf{F}_n\mathbf{P}_n\mathbf{H}_n^T\mathbf{K}_n^T - \mathbf{K}_n\mathbf{H}_n\mathbf{P}_n\mathbf{F}_n^T \\ &+ \mathbf{K}_n\mathbf{H}_n\mathbf{P}_n\mathbf{H}_n^T\mathbf{K}_n^T + \mathbf{K}_n\mathbf{R}_n\mathbf{K}_n^T + \mathbf{Q}_n.\end{aligned} \tag{87}$$

To determine $\mathbf{K}_n$, it is necessary to minimize the mean square error

$$\varepsilon = \mathrm{Tr}\ (\mathbf{P}_{n+1}). \tag{88}$$

Note that the trace is a scalar quantity and therefore ε is a scalar function of the elements $(\mathbf{K}_n)_{ij}$ of the matrix $\mathbf{K}_n$. Furthermore, since $\mathbf{P}_n$ is independent of $\mathbf{K}_n$, the minimum can be obtained by straightforward differentiation. Thus,

$$\frac{\partial\varepsilon}{\partial K_{ij}} = 0, \tag{89}$$

for $i = 1, \ldots, m$, and $j = 1, \ldots, n$. This may be written more compactly in the matrix form

$$\frac{\partial\varepsilon}{\partial\mathbf{K}} = 0, \tag{90}$$

where the indicated derivative is a matrix with the i, j element given by equation (89). The derivative can be computed easily with the use of the following identities:

$$\frac{\partial}{\partial\mathbf{A}}\mathrm{Tr}\ (\mathbf{AB}) = \mathbf{B}^T, \tag{91}$$

$$\frac{\partial}{\partial\mathbf{A}}\mathrm{Tr}\ (\mathbf{BA}^T) = \mathbf{B}, \tag{92}$$

$$\frac{\partial}{\partial\mathbf{A}}\mathrm{Tr}\ (\mathbf{ABA}^T) = 2\mathbf{AB}. \tag{93}$$

These are easily proved. (For example, write out the matrix products for the case when all the matrices are 2×2 and then take the derivative of the trace.) Using these formulas to differentiate the trace of equation (87) it is a simple matter to obtain

$$\frac{\partial \varepsilon}{\partial \mathbf{K}_n} = -\mathbf{F}_n\mathbf{P}_n\mathbf{F}_n^T - (\mathbf{H}_n\mathbf{P}_n\mathbf{F}_n^T)^T + 2\mathbf{K}_n\mathbf{H}_n\mathbf{P}_n\mathbf{H}_n^T + 2\mathbf{K}_n\mathbf{R}_n = 0. \tag{94}$$

Rearranging terms, this becomes

$$\mathbf{K}_n(\mathbf{H}_n\mathbf{P}_n\mathbf{H}_n^T + \mathbf{R}_n) = \mathbf{F}_n\mathbf{P}_n\mathbf{H}_n^T. \tag{95}$$

Note that since both $\mathbf{P}_n$ and $\mathbf{R}_n$ are symmetric matrices, the term

$$(\mathbf{H}_n\mathbf{P}_n\mathbf{H}_n^T + \mathbf{R}_n) \tag{96}$$

is also symmetric. Furthermore, since both $\mathbf{P}_n$ and $\mathbf{R}_n$ are positive definite, this implies that the term (96) is also positive definite. Consequently, all its eigenvalues are positive which implies that it is nonsingular. Therefore, the inverse exists and the solution of (95) is given by

$$\mathbf{K}_n = \mathbf{F}_n\mathbf{P}_n\mathbf{H}_n^T(\mathbf{H}_n\mathbf{P}_n\mathbf{H}_n^T + \mathbf{R}_n)^{-1}. \tag{97}$$

This is the main result. To be rigorous, it must be verified that this is indeed a minimum and not a maximum. This detail is omitted for the sake of brevity.

To complete the derivation the result (97) for $\mathbf{K}_n$ may be substituted back into equation (87) to find a simple form for $\mathbf{P}_{n+1}$. Substituting (97) into (87) gives $\mathbf{P}_{n+1} =$

$$\begin{aligned}
& \mathbf{F}_n\mathbf{P}_n\mathbf{F}_n^T - \mathbf{F}_n\mathbf{P}_n\mathbf{H}_n^T[(\mathbf{H}_n\mathbf{P}_n\mathbf{H}_n^T + \mathbf{R}_n)^{-1}]^T\mathbf{H}_n\mathbf{P}_n\mathbf{F}_n^T \\
- \; & \mathbf{F}_n\mathbf{P}_n\mathbf{H}_n^T(\mathbf{H}_n\mathbf{P}_n\mathbf{H}_n^T + \mathbf{R}_n)^{-1}\mathbf{H}_n\mathbf{P}_n\mathbf{F}_n^T \\
+ \; & \mathbf{F}_n\mathbf{P}_n\mathbf{H}_n^T(\mathbf{H}_n\mathbf{P}_n\mathbf{H}_n^T + \mathbf{R}_n)^{-1}\mathbf{H}_n\mathbf{P}_n\mathbf{H}_n^T[(\mathbf{H}_n\mathbf{P}_n\mathbf{H}_n^T + \mathbf{R}_n)^{-1}]^T\mathbf{H}_n\mathbf{P}_n\mathbf{F}_n^T \\
+ \; & \mathbf{F}_n\mathbf{P}_n\mathbf{H}_n^T(\mathbf{H}_n\mathbf{P}_n\mathbf{H}_n^T + \mathbf{R}_n)^{-1}\mathbf{R}_n[(\mathbf{H}_n\mathbf{P}_n\mathbf{H}_n^T + \mathbf{R}_n)^{-1}]^T\mathbf{H}_n\mathbf{P}_n\mathbf{F}_n^T \\
+ \; & \mathbf{Q}_n.
\end{aligned} \tag{98}$$

Since the term (96) is symmetric, the second to last two terms in (98) can be combined to give

$$\mathbf{P}_{n+1} = \mathbf{F}_n\mathbf{P}_n\mathbf{F}_n^T - \mathbf{F}_n\mathbf{P}_n\mathbf{H}_n^T(\mathbf{H}_n\mathbf{P}_n\mathbf{H}_n^T + \mathbf{R}_n)^{-1}\mathbf{H}_n\mathbf{P}_n\mathbf{F}_n^T + \mathbf{Q}_n. \tag{99}$$

This may be further simplified using (97). Hence,

$$\mathbf{P}_{n+1} = (\mathbf{F}_n - \mathbf{K}_n\mathbf{H}_n)\mathbf{P}_n\mathbf{F}_n^T + \mathbf{Q}_n. \tag{100}$$

This is the desired result.

This completes the derivation of the Kalman filter in n-dimensions. In summary, the Kalman filter equations are given by

$$\begin{aligned}
\hat{\mathbf{x}}_{n+1} &= \mathbf{F}_n\hat{\mathbf{x}}_n + \mathbf{K}_n(\mathbf{z}_n - \mathbf{H}_n\hat{\mathbf{x}}_n), && (101) \\
\mathbf{P}_{n+1} &= (\mathbf{F}_n - \mathbf{K}_n\mathbf{H}_n)\mathbf{P}_n\mathbf{F}_n^T + \mathbf{Q}_n, && (102) \\
\mathbf{K}_n &= \mathbf{F}_n\mathbf{P}_n\mathbf{H}_n^T(\mathbf{H}_n\mathbf{P}_n\mathbf{H}_n^T + \mathbf{R}_n)^{-1}, && (103)
\end{aligned}$$

where $n = 0, 1, 2, \ldots$. The initial conditions are given by

$$\hat{\mathbf{x}}_0 = \bar{\mathbf{x}}_0, \tag{104}$$

$$\mathbf{P}_0 = \bar{\mathbf{P}}_0. \tag{105}$$

One implementation of the Kalman filter algorithm is as follows:

1. Set $n = 0$,
2. Initialize $\hat{\mathbf{x}}_0 = \bar{\mathbf{x}}_0$,
3. Initialize $\mathbf{P}_0 = \bar{\mathbf{P}}_0$,
4. Calculate $\mathbf{K}_n$ from equation (103),
5. Calculate $\hat{\mathbf{x}}_{n+1}$ from (101),
6. Calculate $\mathbf{P}_{n+1}$ from (102),
7. Increment n and goto step 4.

If desired, the sequences $\mathbf{K}_n$ and $\mathbf{P}_n$ may be precomputed before running the algorithm.

6 Standard Kalman Filter Equations

In some instances, it is necessary to estimate the current state $\mathbf{x}_n$ based on measurements up to and including $\mathbf{z}_n$. In the last section, the measurements $\mathbf{z}_0, \mathbf{z}_1, \ldots, \mathbf{z}_n$ were used to estimate the state $\mathbf{x}_{n+1}$. In this section, the measurements $\mathbf{z}_0, \mathbf{z}_1, \ldots, \mathbf{z}_n$ will be used to estimate the state $\mathbf{x}_n$. Assuming a recursive form of solution, the calculation proceeds in exactly the same way as in the previous section. However, it must be understood that the quantities $\hat{\mathbf{x}}_n$, $\mathbf{P}_n$, and $\mathbf{K}_n$ which appear in this section are not the same as those in section 5. Relationships between the two solutions will be obtained in section 8. For the time being, the results of this section are completely independent of those of the last section. A separate solution for the 1-dimensional problem is omitted since it is a special case of the more general $n-$dimensional solution. The interested reader may derive the solution to the 1-dimensional case as an excersize.

A word about terminology. The term "standard Kalman filter" is not a part of the Kalman filtering literature nor is it commonly used by workers in this field. It has been chosen by the author because the recursive solution to the linear prediction problem of all orders may be obtained directly from the solution of the "standard" Kalman filter equations. In section 8, for example, it will be shown how the 1-step prediction may be obtained from the standard solution.

Problem Statement: A linear system obeys the state space equations (67) together with the measurement equation (68). Given the measurements $\mathbf{z}_0, \mathbf{z}_1, \ldots, \mathbf{z}_n$, find the unbiased least squares estimate $\hat{\mathbf{x}}_n$ of the state $\mathbf{x}_n$. Assume that the estimate has the linear recursive form

$$\hat{\mathbf{x}}_n = \mathbf{K}_n \mathbf{z}_n + \mathbf{K}'_n \hat{\mathbf{x}}_{n-1}, \tag{106}$$

where $n = 1, 2, \ldots$.

Problem Solution: To begin, use (106) to write

$$\hat{\mathbf{x}}_n - \mathbf{x}_n = (\mathbf{K}_n \mathbf{z}_n + \mathbf{K}'_n \hat{\mathbf{x}}_{n-1}) - \mathbf{x}_n. \tag{107}$$

Substituting for $\mathbf{x}_n$ and $\mathbf{z}_n$ from the state equation (67) and the measurement equation (68), this becomes

$$\hat{\mathbf{x}}_n - \mathbf{x}_n = (\mathbf{K}_n \mathbf{H}_n - 1)\mathbf{F}_{n-1}\mathbf{x}_{n-1} + \mathbf{K}'_n \hat{\mathbf{x}}_{n-1} + (\mathbf{K}_n \mathbf{H}_n - 1)\mathbf{w}_{n-1} + \mathbf{K}_n \mathbf{v}_n. \tag{108}$$

Taking the expectation of this equation, the requirement that the estimate be unbiased implies

$$E[\hat{\mathbf{x}}_n - \mathbf{x}_n] = (\mathbf{K}_n \mathbf{H}_n - 1)\mathbf{F}_{n-1} E[\mathbf{x}_{n-1}] + \mathbf{K}'_n E[\hat{\mathbf{x}}_{n-1}] = 0. \tag{109}$$

Therefore, since $E[\hat{\mathbf{x}}_{n-1} - \mathbf{x}_{n-1}] = 0$, this yields

$$\mathbf{K}'_n = (1 - \mathbf{K}_n \mathbf{H}_n)\mathbf{F}_{n-1}. \tag{110}$$

Substituting this back into equation (108), it follows that

$$(\hat{\mathbf{x}}_n - \mathbf{x}_n) = (1 - \mathbf{K}_n \mathbf{H}_n)\mathbf{F}_{n-1}(\hat{\mathbf{x}}_{n-1} - \mathbf{x}_{n-1}) + \mathbf{K}_n \mathbf{v}_n + (\mathbf{K}_n \mathbf{H}_n - 1)\mathbf{w}_{n-1}. \tag{111}$$

Next, the factor $\mathbf{K}_n$ is determined by minimizing the mean square error

$$\varepsilon = \|\hat{\mathbf{x}}_n - \mathbf{x}_n\|^2. \tag{112}$$

To simplify the process of taking derivatives, write this in the form of a trace, that is,

$$\varepsilon = \mathrm{Tr}\, E[(\hat{\mathbf{x}}_n - \mathbf{x}_n)(\hat{\mathbf{x}}_n - \mathbf{x}_n)^T]. \tag{113}$$

Using equation (111) together with the linearity of the expectation, this yields

$$\begin{aligned} &E[(\hat{\mathbf{x}}_n - \mathbf{x}_n)(\hat{\mathbf{x}}_n - \mathbf{x}_n)^T] = \\ &(1 - \mathbf{K}_n \mathbf{H}_n)\mathbf{F}_{n-1} E[(\hat{\mathbf{x}}_{n-1} - \mathbf{x}_{n-1})(\hat{\mathbf{x}}_{n-1} - \mathbf{x}_{n-1})^T]\mathbf{F}^T_{n-1}(1 - \mathbf{K}_n \mathbf{H}_n)^T \\ &\quad + \mathbf{K}_n E[\mathbf{v}_n \mathbf{v}_n^T]\mathbf{K}_n^T + (\mathbf{K}_n \mathbf{H}_n - 1)E[\mathbf{w}_{n-1}\mathbf{w}^T_{n-1}](\mathbf{K}_n \mathbf{H}_n - 1)^T. \end{aligned} \tag{114}$$

Defining the quantity $\mathbf{P}_n$ by

$$\mathbf{P}_n = E[(\hat{\mathbf{x}}_n - \mathbf{x}_n)(\hat{\mathbf{x}}_n - \mathbf{x}_n)^T], \tag{115}$$

the last equation becomes

$$\begin{aligned}\mathbf{P}_n &= (1-\mathbf{K}_n\mathbf{H}_n)\mathbf{F}_{n-1}\mathbf{P}_{n-1}\mathbf{F}_{n-1}^T(1-\mathbf{K}_n\mathbf{H}_n)^T \\ &+ \mathbf{K}_n\mathbf{R}_n\mathbf{K}_n^T + (1-\mathbf{K}_n\mathbf{H}_n)\mathbf{Q}_{n-1}(1-\mathbf{K}_n\mathbf{H}_n)^T.\end{aligned} \tag{116}$$

Note that $\mathbf{P}_{n-1}$ is independent of $\mathbf{K}_n$, and therefore is a constant for purposes of differentiation. Now, the mean square error is given by

$$\varepsilon = \mathrm{Tr}\ (\mathbf{P}_n), \tag{117}$$

and this will be a minimum if

$$\frac{\partial}{\partial \mathbf{K}_n}\mathrm{Tr}\ (\mathbf{P}_n) = 0, \tag{118}$$

where the derivative with respect to a matrix is defined componentwise as in (89). Expanding the product in equation (116) yields

$$\begin{aligned}\mathbf{P}_n &= \mathbf{F}_{n-1}\mathbf{P}_{n-1}\mathbf{F}_{n-1}^T - \mathbf{F}_{n-1}\mathbf{P}_{n-1}\mathbf{F}_{n-1}^T\mathbf{H}_n^T\mathbf{K}_n^T - \mathbf{K}_n\mathbf{H}_n\mathbf{F}_{n-1}\mathbf{P}_{n-1}\mathbf{F}_{n-1}^T \\ &+ \mathbf{K}_n\mathbf{H}_n\mathbf{F}_{n-1}\mathbf{P}_{n-1}\mathbf{F}_{n-1}^T\mathbf{H}_n^T\mathbf{K}_n^T + \mathbf{K}_n\mathbf{R}_n\mathbf{K}_n^T + \mathbf{Q}_{n-1} \\ &- \mathbf{Q}_{n-1}\mathbf{H}_n^T\mathbf{K}_n^T - \mathbf{K}_n\mathbf{H}_n\mathbf{Q}_{n-1} + \mathbf{K}_n\mathbf{H}_n\mathbf{Q}_{n-1}\mathbf{H}_n^T\mathbf{K}_n^T.\end{aligned} \tag{119}$$

Taking the trace of this equation and differentiating using the identities (91), (92), and (93), it follows that

$$\begin{aligned}\frac{\partial}{\partial \mathbf{K}_n}\mathrm{Tr}\ (\mathbf{P}_n) &= -2\mathbf{F}_{n-1}\mathbf{P}_{n-1}\mathbf{F}_{n-1}^T\mathbf{H}_n^T + 2\mathbf{K}_n\mathbf{H}_n\mathbf{F}_{n-1}\mathbf{P}_{n-1}\mathbf{F}_{n-1}^T\mathbf{H}_n^T \\ &+ 2\mathbf{K}_n\mathbf{R}_n - 2\mathbf{Q}_{n-1}\mathbf{H}_n^T + 2\mathbf{K}_n\mathbf{H}_n\mathbf{Q}_{n-1}\mathbf{H}_n^T.\end{aligned} \tag{120}$$

The fact that $\mathbf{R}_n$ and $\mathbf{Q}_n$ are both symmetric has also been used. Equating this to zero and collecting terms yields

$$\begin{aligned}&\mathbf{K}_n\left[\mathbf{H}_n\mathbf{F}_{n-1}\mathbf{P}_{n-1}\mathbf{F}_{n-1}^T\mathbf{H}_n^T + \mathbf{R}_n + \mathbf{H}_n\mathbf{Q}_{n-1}\mathbf{H}_n^T\right] = \\ &\qquad\qquad \left[\mathbf{F}_{n-1}\mathbf{P}_{n-1}\mathbf{F}_{n-1}^T\mathbf{H}_n^T + \mathbf{Q}_{n-1}\mathbf{H}_n^T\right].\end{aligned} \tag{121}$$

This expression can be made more manageable by making the definition

$$\mathbf{M}_n = \mathbf{F}_n\mathbf{P}_n\mathbf{F}_n^T + \mathbf{Q}_n. \tag{122}$$

Hence, equation (121) may be written

$$\mathbf{K}_n\left[\mathbf{H}_n\mathbf{M}_{n-1}\mathbf{H}_n^T + \mathbf{R}_n\right] = \mathbf{M}_{n-1}\mathbf{H}_n^T. \tag{123}$$

If the inverse exists, then the solution for $\mathbf{K}_n$ takes the form

$$\mathbf{K}_n = \mathbf{M}_{n-1}\mathbf{H}_n^T\left[\mathbf{H}_n\mathbf{M}_{n-1}\mathbf{H}_n^T + \mathbf{R}_n\right]^{-1}. \tag{124}$$

But it is easy to see that the inverse does exist since $\mathbf{R}_n$, $\mathbf{Q}_n$, and $\mathbf{P}_n$, are all symmetric, positive definite matrices, and all positive definite matrices are nonsingular. Therefore,

(124) is the final solution for $\mathbf{K}_n$. The proof that (124) is indeed a minimum and not a maximum will be omitted.

The expression (119) for $\mathbf{P}_n$ can be simplified. First, write equation (119) in the form

$$\begin{aligned} \mathbf{P}_n &= \mathbf{K}_n(\mathbf{H}_n\mathbf{M}_{n-1}\mathbf{H}_n^T + \mathbf{R}_n)\mathbf{K}_n^T \\ &- \mathbf{K}_n\mathbf{H}_n\mathbf{M}_{n-1} - \mathbf{M}_{n-1}\mathbf{H}_n^T\mathbf{K}_n^T + \mathbf{M}_{n-1}. \end{aligned} \tag{125}$$

Substituting the result (124) for $\mathbf{K}_n$ in the first term, but not for $\mathbf{K}_n^T$, the first and third terms cancel, leaving the result

$$\mathbf{P}_n = (1 - \mathbf{K}_n\mathbf{H}_n)\mathbf{M}_{n-1}. \tag{126}$$

This is the desired form.

In summary, the standard Kalman filter equations are given by

$$\hat{\mathbf{x}}_n = \mathbf{K}_n\mathbf{z}_n + (1 - \mathbf{K}_n\mathbf{H}_n)\mathbf{F}_{n-1}\hat{\mathbf{x}}_{n-1}, \tag{127}$$

$$\mathbf{K}_n = \mathbf{M}_{n-1}\mathbf{H}_n^T[\mathbf{H}_n\mathbf{M}_{n-1}\mathbf{H}_n^T + \mathbf{R}_n]^{-1}, \tag{128}$$

$$\mathbf{M}_{n-1} = \mathbf{F}_{n-1}\mathbf{P}_{n-1}\mathbf{F}_{n-1}^T + \mathbf{Q}_{n-1}, \tag{129}$$

$$\mathbf{P}_n = (1 - \mathbf{K}_n\mathbf{H}_n)\mathbf{M}_{n-1}, \tag{130}$$

where $n = 1, 2, \ldots$ In order to initialize the Kalman filter algorithm, it is necessary to define initial conditions for $\hat{\mathbf{x}}_0$, $\mathbf{K}_0$, and $\mathbf{P}_0$. These initial conditions will be derived in the next section.

7 Initial Conditions for the Standard Kalman Filter

To determine the initial conditions, the problem is to find the estimate $\hat{\mathbf{x}}_0$ of the initial state $\mathbf{x}_0$ based on the one measurement $\mathbf{z}_0$. The solution to this problem is straightforward and may be obtained by the methods used previously. Once the initial conditions are determined, a summary of the resulting Kalman filter algorithm will be presented.

Problem Statement: Let

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{v}, \tag{131}$$

where $\mathbf{x}$ is a random vector with mean $\bar{\mathbf{x}}$ and covariance matrix $\bar{\mathbf{P}}$, and $\mathbf{v}$ is a zero mean random vector which is uncorrelated with $\mathbf{x}$. Find the unbiased, linear, least squares estimate $\hat{\mathbf{x}}$ of the vector $\mathbf{x}$ based on the measurement $\mathbf{z}$. Assume that the estimate has the form

$$\hat{\mathbf{x}} = \mathbf{K}\mathbf{z} + \mathbf{K}'\bar{\mathbf{x}}. \tag{132}$$

This problem bears a close resemblance to the least squares problem of section 2. The two problems, however, are fundamentally different. The difference is that in equation (8), section 2, the vector $\mathbf{x}$ was a constant, whereas in equation (131) the vector $\mathbf{x}$ is a random

variable.

Problem Solution: To begin, use equations (131) and (132) to write

$$(\hat{\mathbf{x}} - \mathbf{x}) = (\mathbf{KH} - 1)\mathbf{x} + \mathbf{K}'\bar{\mathbf{x}} + \mathbf{Kv}. \tag{133}$$

Taking the expectation and using the requirement that $\hat{\mathbf{x}}$ is unbiased, it follows that

$$\mathbf{K}' = (1 - \mathbf{KH}). \tag{134}$$

bstituting this back into (133) yields

$$(\hat{\mathbf{x}} - \mathbf{x}) = (\mathbf{KH} - 1)(\mathbf{x} - \bar{\mathbf{x}}) + \mathbf{Kv}. \tag{135}$$

To determine $\mathbf{K}$ it is necessary to minimize the mean square error

$$\varepsilon = \mathrm{Tr}\ E[(\hat{\mathbf{x}} - \mathbf{x})(\hat{\mathbf{x}} - \mathbf{x})^T]. \tag{136}$$

Using equation (135) together with the linearity of the expectation, it follows that

$$\begin{aligned} E[(\hat{\mathbf{x}} - \mathbf{x})(\hat{\mathbf{x}} - \mathbf{x})^T] &= (\mathbf{KH} - 1)E[(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^T](\mathbf{KH} - 1)^T \\ &+ \mathbf{KRK}^T. \end{aligned} \tag{137}$$

Making the definition

$$\mathbf{P} = E[(\hat{\mathbf{x}} - \mathbf{x})(\hat{\mathbf{x}} - \mathbf{x})^T], \tag{138}$$

and using the given covariance matrix

$$\bar{\mathbf{P}} = E[(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^T], \tag{139}$$

this equation becomes

$$\mathbf{P} = (\mathbf{KH} - 1)\bar{\mathbf{P}}(\mathbf{KH} - 1)^T + \mathbf{KRK}^T. \tag{140}$$

Then, expanding the product, this may be written

$$\mathbf{P} = \mathbf{KH\bar{P}H}^T\mathbf{K}^T - \mathbf{KH\bar{P}} - \bar{\mathbf{P}}\mathbf{H}^T\mathbf{K}^T + \bar{\mathbf{P}} + \mathbf{KRK}^T. \tag{141}$$

Taking the trace of this equation and then the derivative with respect to $\mathbf{K}$ yields

$$\frac{\partial \varepsilon}{\partial \mathbf{K}} = 2\mathbf{KH\bar{P}H}^T - 2\bar{\mathbf{P}}\mathbf{H}^T + 2\mathbf{KR} = 0. \tag{142}$$

And, finally, solving this equation for $\mathbf{K}$,

$$\mathbf{K} = \bar{\mathbf{P}}\mathbf{H}^T(\mathbf{H\bar{P}H}^T + \mathbf{R})^{-1}. \tag{143}$$

This is a familiar result. Clearly, the inverse in (143) exists since $\bar{\mathbf{P}}$ and $\mathbf{R}$ are both positive definite.

To find a simpified form for $\mathbf{P}$, first write equation (140) in the form

$$\mathbf{P} = \mathbf{K}(\mathbf{H}\bar{\mathbf{P}}\mathbf{H}^T + \mathbf{R})\mathbf{K}^T - \mathbf{K}\mathbf{H}\bar{\mathbf{P}} - \bar{\mathbf{P}}\mathbf{H}^T\mathbf{K}^T + \bar{\mathbf{P}}. \tag{144}$$

Then, substituting (143) for $\mathbf{K}$ in the first term but not for $\mathbf{K}^T$, the first and third terms cancel, leaving

$$\mathbf{P} = (1 - \mathbf{K}\mathbf{H})\bar{\mathbf{P}}. \tag{145}$$

This is the desired form.

Having solved the above estimation problem, the initial conditions for the standard Kalman filter equations may be obtained by simply putting a subscript '0' on each of the equations (132), (143), and (145). Thus,

$$\begin{aligned} \hat{\mathbf{x}}_0 &= \bar{\mathbf{x}}_0 + \mathbf{K}_0(\mathbf{z}_0 - \mathbf{H}_0\bar{\mathbf{x}}_0), & (146)\\ \mathbf{P}_0 &= (1 - \mathbf{K}_0\mathbf{H}_0)\bar{\mathbf{P}}_0, & (147)\\ \mathbf{K}_0 &= \bar{\mathbf{P}}_0\mathbf{H}_0^T(\mathbf{H}_0\bar{\mathbf{P}}_0\mathbf{H}_0^T + \mathbf{R}_0)^{-1}. & (148) \end{aligned}$$

An outline of the complete Kalman filter algorithm is as follows:

1. Initialize $\mathbf{K}_0$ using (148),
2. Initialize $\hat{\mathbf{x}}_0$ using (146),
3. Initialize $\mathbf{P}_0$ using (147),
4. Set $n = 1$,
5. Calculate $\mathbf{M}_{n-1}$ from (129),
6. Calculate $\mathbf{K}_n$ from equation (128),
7. Calculate the state estimate $\hat{\mathbf{x}}_n$ from (127),
8. Calculate $\mathbf{P}_n$ from (130),
9. Increment n and goto step 5.

8 Prediction–Correction Formulation

In this section, a relation will be found between the state estimate $\hat{\mathbf{x}}_n$ of section 6, and the state prediction $\hat{\mathbf{x}}_{n+1}$ of section 5. Using this result, the preveously separate results of sections 5 and 6 may be combined into one unified set of equations for both the current state estimate $\hat{\mathbf{x}}_n$ (section 6) and the state prediction $\hat{\mathbf{x}}_n$ (section 5). The unified set of Kalman filter equations consists of the standard Kalman filter equations of section 6, together with a new equation for the state prediction $\hat{\mathbf{x}}_{n+1}$ in terms of the current state estimate $\hat{\mathbf{x}}_n$. This new equation will be derived in this section. First, to prevent confusion, it is necessary to introduce some new notation.

Let $\hat{\mathbf{x}}(n|n)$ be the estimate of the state $\mathbf{x}_n$ based on the measurements $\mathbf{z}_0, \ldots, \mathbf{z}_n$, and let $\hat{\mathbf{x}}(n+1|n)$ be the estimate of the state $\mathbf{x}_{n+1}$ based on the measurements $\mathbf{z}_0, \ldots, \mathbf{z}_n$. Then $\hat{\mathbf{x}}(n|n)$ is the solution to the Kalman filter equations (127), (128), (129), and (130), which were derived in section 6. And $\hat{\mathbf{x}}(n+1|n)$ is the solution to the Kalman predictor equations (103), (104), and (105), which were derived in section 5. It will now be shown that the one-step prediction $\hat{\mathbf{x}}(n+1|n)$ may be obtained from the current state estimate $\hat{\mathbf{x}}(n|n)$ by the simple relation

$$\hat{\mathbf{x}}(n+1|n) = \mathbf{F}_n \hat{\mathbf{x}}(n|n). \tag{149}$$

A simple, heuristic proof will be given here. A more detailed proof is given in the appendix. Begin with the fact that, by definition, the estimate $\hat{\mathbf{x}}(n|n)$ and the estimate $\hat{\mathbf{x}}(n+1|n)$ are both linear combinations of the measurements $\mathbf{z}_0, \mathbf{z}_1, \ldots, \mathbf{z}_n$. Therefore, it is reasonable to assume that they are related by a linear transformation of the form

$$\hat{\mathbf{x}}(n+1|n) = \mathbf{A}_n \hat{\mathbf{x}}(n|n), \tag{150}$$

where $\mathbf{A}_n$ is a square matrix. From the state equation (67) it follows that

$$\hat{\mathbf{x}}(n+1|n) - \mathbf{x}_{n+1} = \mathbf{A}_n \hat{\mathbf{x}}(n|n) - (\mathbf{F}_n \mathbf{x}_n + \mathbf{w}_n). \tag{151}$$

Taking the expectation of this equation yields

$$E[\hat{\mathbf{x}}(n+1|n) - \mathbf{x}_{n+1}] = \mathbf{A}_n E[\hat{\mathbf{x}}(n|n)] - \mathbf{F}_n E[\mathbf{x}_n] = 0. \tag{152}$$

But, since $\hat{\mathbf{x}}(n|n)$ is an unbiased estimate of $\mathbf{x}_n$,

$$E[\hat{\mathbf{x}}(n|n)] = E[\mathbf{x}_n]. \tag{153}$$

Therefore, equation (152) implies

$$\mathbf{A}_n = \mathbf{F}_n. \tag{154}$$

This is the desired result.

By adding equation (149) to the standard Kalman filter equations (127)-(130), these may be rewritten in the following form:

$$\begin{aligned}
\hat{\mathbf{x}}(n|n) &= \hat{\mathbf{x}}(n+1|n) + \mathbf{K}_n[\mathbf{z}_n - \mathbf{H}_n \hat{\mathbf{x}}(n|n-1)], && (155)\\
\hat{\mathbf{x}}(n+1|n) &= \mathbf{F}_n \hat{\mathbf{x}}(n|n), && (156)\\
\mathbf{K}_n &= \mathbf{M}_{n-1}\mathbf{H}_n^T[\mathbf{H}_n \mathbf{M}_{n-1}\mathbf{H}_n^T + \mathbf{R}_n]^{-1}, && (157)\\
\mathbf{M}_{n-1} &= \mathbf{F}_{n-1}\mathbf{P}_{n-1}\mathbf{F}_{n-1}^T + \mathbf{Q}_{n-1}, && (158)\\
\mathbf{P}_n &= (1 - \mathbf{K}_n \mathbf{H}_n)\mathbf{M}_{n-1}, && (159)
\end{aligned}$$

where $n = 0, 1, 2, \ldots$. The initial conditions are now given by

$$\begin{aligned}
\hat{\mathbf{x}}(0|-1) &= \bar{\mathbf{x}}_0, && (160)\\
\mathbf{M}_{-1} &= \bar{\mathbf{P}}_0. && (161)
\end{aligned}$$

When written in this way, the Kalman filter is a combination of the state update equation (155) and the state prediction equation (156). These are commonly refered to as the prediction-correction formulas. This is a convenient form for most applications. A direct implementation of the Kalman filter algorithm is as follows:

1. Initialize $\hat{\mathbf{x}}(0|-1) = \bar{\mathbf{x}}_0$,

2. Initialize $\mathbf{M}_{-1} = \bar{\mathbf{P}}_0$,

3. Set $n = 0$,

4. Calculate $\mathbf{K}_n$ from (157),

5. Calculate $\hat{\mathbf{x}}(n|n)$ from (155),

6. Calculate $\hat{\mathbf{x}}(n+1|n)$ from (156),

7. Calculate $\mathbf{M}_{n-1}$ from (158),

8. Calculate $\mathbf{P}_n$ from (159),

9. Increment n and goto step 4.

Having presented the mathematical theory of the Kalman filter algorithms, it is important to note that there are certain pitfalls involved with the computational aspects of these algorithms. These problems are inevitable due to the finite precision of digital computers. More information about the practical implementation of the Kalman filter can be found in the literature. [14]–[16]

9 Summary

A rigorous derivation of the Kalman filter equations has been presented which is appropriate for an introductory short course on Kalman filter theory. The way it has been defined, the Kalman filter gives the optimum unbiased, linear, least squares estimate of the state vector based on all past and current measurements. At this point, it is strongly recommended that the reader study a few applications of the Kalman filter to see how the filter is used in practical problems. One application is in the radar tracking of airborne targets. This application is well described in some of the original papers.[17]–[19] The range of applications of the Kalman filter is vast and varied. More applications can be found in the IEEE reprint series *Kalman Filtering: Theory and Application*.[14]

References

[1] Kalman, R.E., "A New Approach to Linear Filtering and Prediction Problems," *Trans. AMSE, Journal of Basic Engineering*, Vol. 82D, pp. 35-45, 1960.

[2] Kalman, R.E., and Bucy, R.S., "New Results in Linear Filtering and Prediction Theory," *Trans. AMSE, Journal of Basic Engineering*, Vol. 83D, pp. 95-108, 1961.

[3] Kalman, R.E., "New Methods and Results in Linear Prediction and Filtering Theory," Technical Report 61-1, Research Institute for Advanced Studies (RIAS), Martin Company, Baltimore, Maryland, 1961.

[4] Kalman, R.E., "New Methods in Wiener Filtering Theory," in Bogdanoff, J.L., and Kozin, F., eds., *Proceedings of the First Symposium on Engineering Applications of Random Function Theory and Probability*, John Wiley and Sons, New York, pp. 270-388, 1963.

[5] Kalman, R.E., Falb, P.L., and Arbib, M.A., *Topics in Mathematical System Theory*, McGraw-Hill, New York, 1969.

[6] Meditch, J.S., *Stochastic Optimal Linear Estimation and Control*, McGraw-Hill, New York, 1969.

[7] Gelb, A., ed., *Applied Optimal Estimation*, M.I.T. Press, Cambridge, Massachusetts, 1974.

[8] Anderson, B.D.O., and Moore, J B., *Optimal Filtering*, Prentice-Hall, Englewood Cliffs, New Jersey, 1979.

[9] Maybeck, P.S., *Stochastic Models, Estimation and Control*, Vol. 1, Academic Press, New York, 1979.

[10] Assefi, T., *Stochastic Processes and Estimation Theory with Applications*, John Wiley, New York, 1979.

[11] Chui, C.K., and Chen, G., *Kalman Filtering with Real-Time Applications*, 2nd ed., Springer-Verlag, Berlin, 1991.

[12] Kailath, T., ed., *Linear Least-Squares Estimation*, Dowden, Hutchinson, & Ross, Stroudsburg, Pennsylvania, 1977.

[13] Yaglom, A.M., *Stationary Random Functions*, Prentice Hall, Englewood Cliffs, New Jersey, 1962.

[14] Sorenson, H.W., ed., *Kalman Filtering: Theory and Application*, IEEE Press, New York, 1985.

[15] Bierman, G.J., *Factorization Methods for Discrete Sequential Estimation*, Academic Press, New York, 1977.

[16] Kaminski, P.G., Bryson, A.E., and Schmidt, S.F., "Discrete Square-Root Filtering: A Survey of Current Techniques," *IEEE Transactions on Automatic Control,* Vol. AC-16, pp. 727-735, 1971.

[17] Singer, R.A., "Estimating Optimal Tracking Filter Performance for Manned Maneuvering Targets," *IEEE Transactions on Aerospace and Electronic Systems,* Vol. AES-6, pp. 473-483, 1970.

[18] Singer, R.A., and Behnke, K., "Real-Time Tracking Filter Evaluation and Selection for Tactical Applications," *IEEE Transactions on Aerospace and Electronic Systems,* Vol. AES-7, pp. 100-110, 1971.

[19] Castella, F.R., and Dunnebacke, F.G., "Analytical Results for the x, y Kalman Tracking Filter," *IEEE Transactions on Aerospace and Electronic Systems,* Vol. AES-10, pp. 891-895, 1974.

Appendix

Proof of Equation 149

The purpose of this appendix is to give an alternate proof of equation (149) in section 8. It is easiest to prove this in the 1-dimensional case first. By the definition of $\hat{x}(n+1|n)$,

$$E[\|\hat{x}(n+1|n) - x_{n+1}\|^2] = \min. \tag{1}$$

In this notation, the minimum is taken with respect to all linear combinations $\hat{x}(n+1|n)$ of the measurements $z_0, \ldots, z_n$. Alternatively, consider

$$\begin{aligned} E[\|F_n\hat{x}(n|n) - x_{n+1}\|^2] &= E[\|F_n\hat{x}(n|n) - F_n x_n - w_n\|^2] & (2)\\ &= E[\|F_n\hat{x}(n|n) - F_n x_n\|^2] + E[\|w_n\|^2], & (3) \end{aligned}$$

where the last inequality follows since both x_n and $\hat{x}(n|n)$ are uncorrelated with w_n. Using the linearity of the expectation this becomes

$$E[\|F_n\hat{x}(n|n) - x_{n+1}\|^2] = F_n^2 E[\|\hat{x}(n|n) - x_n\|^2] + E[\|w_n\|^2]. \tag{4}$$

But, by the definition of $\hat{x}(n|n)$,

$$E[\|\hat{x}(n|n) - x_n\|^2] = \min. \tag{5}$$

Consequently, equation (4) implies

$$E[\|F_n\hat{x}(n|n) - x_{n+1}\|] = \min. \tag{6}$$

Now, recall that $\hat{x}(n|n)$ and $\hat{x}(n+1|n)$ are both linear combinations of the measurements $z_0, z_1, \ldots, z_n$. Therefore, if the linear combination that minimizes (1) is unique, then equations (1) and (6) together imply

$$\hat{x}(n+1|n) = F_n\hat{x}(n|n). \tag{7}$$

This completes the proof in the 1-dimensional case.

The proof in the n-dimensional case is similar. By the definition of the estimate $\hat{\mathbf{x}}(n+1|n)$,

$$E[\|\hat{\mathbf{x}}(n+1|n) - \mathbf{x}_{n+1}\|^2] = \min. \tag{8}$$

On the other hand,

$$\begin{aligned} E[\|\mathbf{F}_n\hat{\mathbf{x}}(n|n) - \mathbf{x}_{n+1}\|^2] &= E[\|\mathbf{F}_n\hat{\mathbf{x}}(n|n) - \mathbf{F}_n\mathbf{x}_n - \mathbf{w}_n\|^2] & (9)\\ &= E[\|\mathbf{F}_n\hat{\mathbf{x}}(n|n) - \mathbf{F}_n\mathbf{x}_n\|^2] + E[\|\mathbf{w}_n\|^2], & (10) \end{aligned}$$

where the second equality follows from the fact that both $\mathbf{x}_n$ and $\hat{\mathbf{x}}(n|n)$ are uncorrelated with $\mathbf{w}_n$. Now, if the first term on the right-hand-side is a minimum, that is, if

$$E[\|\mathbf{F}_n\hat{\mathbf{x}}(n|n) - \mathbf{F}_n\mathbf{x}_n\|^2] = \min, \tag{11}$$

then this implies

$$E[\|\mathbf{F}_n\hat{\mathbf{x}}(n|n) - \mathbf{x}_{n+1}\|^2] = \min. \tag{12}$$

Now, observe that both $\mathbf{F}_n\hat{\mathbf{x}}(n|n)$ and $\hat{\mathbf{x}}(n+1|n)$ are linear combinations (really multilinear combinations) of the measurements $\mathbf{z}_0, \mathbf{z}_1, \ldots, \mathbf{z}_n$, that is,

$$\begin{aligned} \hat{\mathbf{x}}(n+1|n) &= \mathbf{A}_0\mathbf{z}_0 + \mathbf{A}_1\mathbf{z}_1 + \ldots + \mathbf{A}_n\mathbf{z}_n, & (13) \\ \mathbf{F}_n\hat{\mathbf{x}}(n|n) &= \mathbf{B}_0\mathbf{z}_0 + \mathbf{B}_1\mathbf{z}_1 + \ldots + \mathbf{B}_n\mathbf{z}_n & (14) \end{aligned}$$

Therefore, if the linear combination (13) which minimizes the mean square error (8) is unique, then equations (8) and (12) will together imply

$$\hat{\mathbf{x}}(n+1|n) = \mathbf{F}_n\hat{\mathbf{x}}(n|n). \tag{15}$$

This is the desired result. Thus, it remains to prove (11). First, note that this may be written

$$E[||\mathbf{F}_n\hat{\mathbf{x}}(n|n) - \mathbf{F}_n\mathbf{x}_n||^2] = E[\mathbf{u}_n^T\mathbf{F}_n^T\mathbf{F}_n\mathbf{u}_n], \tag{16}$$

where

$$\mathbf{u}_n = \hat{\mathbf{x}}(n|n) - \mathbf{x}_n. \tag{17}$$

Setting

$$\mathbf{A} = \mathbf{F}_n^T\mathbf{F}_n, \tag{18}$$

then clearly $\mathbf{A}$ is symmetric, and if $\mathbf{F}_n$ is nonsingular, then $\mathbf{A}$ is positive definite. To prove the desired result, it will be shown that

$$E[\mathbf{u}_n^T\mathbf{A}\mathbf{u}_n] = \min, \tag{19}$$

if

$$E[\mathbf{u}^T\mathbf{u}] = \min, \tag{20}$$

where $\mathbf{A}$ is any symmetric positive definite matrix. Since $\mathbf{A}$ is symmetric, it is orthogonally diagonalizable. Thus, there exists a matrix $\mathbf{S}$ such that

$$\mathbf{S}^T\mathbf{A}\mathbf{S} = \mathbf{D}, \tag{21}$$

where $\mathbf{D}$ is the diagonal matrix

$$\mathbf{D} = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}, \tag{22}$$

and $\lambda_1, \ldots, \lambda_n$, are the eigenvalues of $\mathbf{A}$. Since $\mathbf{A}$ is positive definite, all its eigenvalues are positive real numbers. Now, one may write

$$\begin{aligned} \mathbf{u}_n^T\mathbf{A}\mathbf{u}_n &= \mathbf{u}^T(\mathbf{S}\mathbf{D}\mathbf{S}^T)\mathbf{u} & (23) \\ &= \mathbf{v}^T\mathbf{D}\mathbf{v} & (24) \\ &= \lambda_1 v_1^2 + \ldots + \lambda_n v_n^2, & (25) \end{aligned}$$

where

$$\mathbf{v} = \mathbf{S}^T\mathbf{u}. \tag{26}$$

Taking the expectation of this equation yields

$$E[\mathbf{u}_n^T\mathbf{A}\mathbf{u}_n] = \lambda_1 E[v_1^2] + \ldots + \lambda_n E[v_n^2]. \tag{27}$$

Since all the eigenvelues are positive, this is a minimum if

$$E[v_1^2] \; = \; \min, \tag{28}$$

$$\vdots$$

$$E[v_n^2] \; = \; \min. \tag{29}$$

Equivalently, this is a minimum if

$$E[\mathbf{v}^T\mathbf{v}] = \min. \tag{30}$$

But, using equation (26),

$$\mathbf{v}^T\mathbf{v} = \mathbf{u}^T\mathbf{S}\mathbf{S}^T\mathbf{u} = \mathbf{u}^T\mathbf{u}. \tag{31}$$

Hence, the condition (26) reduces to

$$E[\mathbf{u}^T\mathbf{u}] = \min. \tag{32}$$

This is the desired result. End of proof.

DISTRIBUTION LIST

Commander
Armament Research, Development and Engineering Center
U.S. Army Tank-automative and Armaments Command
ATTN: AMSTA-AR-IMC (3)
SMCAR-GCL
SMCAR-FSP-E, J. Podesta (10)
Picatinny Arsenal, NJ 07806-5000

Administrator
Defense Technical Information Center
ATTN: Accessions Division (12)
Cameron Station
Alexandria, VA 22304-6145

Director
U.S. Army Materiel Systems Analysis Activity
ATTN: AMXSY-MP
Aberdeen Proving Ground, MD 21005-5066

Commander
Chemical/Biological Defense Agency
U.S. Army Armament, Munitions and Chemical Command
ATTN: AMSCB-CII, Library
Aberdeen Proving Ground, MD 21010-5423

Director
U.S. Army Edgewood Research, Development and Engineering Center
ATTN: SCBRD-RTB (Aerodynamics Technology Team)
Aberdeen Proving Ground, MD 21010-5423

Director
U.S. Army Research Laboratory
ATTN: AMSRL-OP-CI-B, Technical Library
Aberdeen Proving Ground, MD 21005-5066

Chief
Benet Weapons Laboratory, CCAC
Armament Research, Development and Engineering Center
U.S. Army Armament, Munitions and Chemical Command
ATTN: SMCAR-CCB-TL
Watervliet, NY 12189-5000

Director
U.S. Army TRADOC Analysis Command-WSMR
ATTN: ATRC-WSS-R
White Sands Missile Range, NM 88002